Министерство образования и науки Российской Федерации
Федеральное государственное бюджетное образовательное учреждение высшего образования
«Московский государственный технический университет имени Н.Э. Баумана (национальный исследовательский университет)»
(МГТУ им. Н.Э. Баумана)

Кафедра «Программное обеспечение ЭВМ и информационные технологии»

ОТЧЕТ
по лабораторная работа №1
по курсу «Математическая статистика»
Вариант №12

Выполнил студент: Мхитарян В.К.
Группа: ИУ7-64
Проверил: Величанский М.А.

2019 г.

1. Формулы для вычисления величин M_{min} , M_{max} , R , $\hat{\mu}, \hat{\sigma}$:

$$M_{min} = X_{(1)}$$

$$M_{min} = X_{(n)}$$

$$R = X_{(n)} - X_{(1)}$$

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} (X_i - \overline{X})^2$$

$$m = [log_2 n] + 2$$

Расположим элементы выборки $\overrightarrow{x}$ в порядке неубывания, тогда последовательность $x_{(1)}, \ldots, x_{(n)}$ называется вариационным рядом выборки $\overrightarrow{x}$.

2. Эмпирической плотностью, отвечающей выборке $\overrightarrow{x}$, называется функция

$$f_n(x) = \begin{cases} \frac{n_i}{n\,\Delta}, & x \in J_i,\ i = \overline{1,m}; \\ 0, & \text{иначе}. \end{cases}, \quad \text{где}$$

Δ - ширина интервала $J = [x_{(1)}, x_{(n)}]$, $\Delta = \dfrac{|J|}{m}$, $m = [log_2 n] + 2$,

$J_i = [X_{(1)} + (i-1)\Delta, x_{(1)} + i\Delta]\ i = \overline{1, m-1}$,

$J_j = [X_{(1)} + (m-1)\Delta, x_{(1)} + m\Delta]$,

n_i - количество элементов выборки, принадлежащих J_i, $i = \overline{1,n}$.

Гистограмма - график эмпирической функции плотности (которая представляет собой кусочно-постоянную функцию на промежутке J).

3. Эмпирическая функция распределения, отвечающая выборке $\overrightarrow{x}_n$, называется функция $F_n(x) = \dfrac{n(x, \overrightarrow{x})}{n}$, где $n(x, \overrightarrow{x})$ - количество элементов выборки, имеющих значение, меньшее x, n - объем выборки.

Если все элементы выборки $\overrightarrow{x}_n$ различны, то функцию можно задать формулой

$$F_n(x) = \begin{cases} 0, x \leq x_{(1)} \\ \frac{i}{n},\ x_{(i)} < x \leq x_{(i+1)} \\ 1, x > x_{(n)} \end{cases}$$

4. Текст программы:

```
function lab1()
   % считать данные из файла
   X = csvread('data.csv');

   X = sort(X);
   n = length(X);

   fprintf("\n%g\n", n);

   % минимальное значение выборки
   minX = min(X);
   fprintf("Min = %4.2f\n", minX);

   % максимальное значение выборки
   maxX = max(X);
   fprintf("Max = %4.2f\n", maxX);

   % размах выборки
   R = maxX - minX;
   fprintf("R = %4.2f\n", R);


   % вычисление оценок мат ожидания и дисперсии
   mu = sum(X)/n;
   fprintf("mu = %4.2f\n", mu);

   % несмещенная оценка дисперсии
   di = sum((X-mu).^2)/(n-1);

   %S2di = disp(X);
   fprintf("sigma^2 = %4.2f\n", di);

   % группировка значений выборки m = [log_2 n] + 2 интервала
   m = floor(log2(n)) + 2;
   fprintf("m = %g\n", m);

   n = length(X);
   m = floor(log2(n)) + 2;
   delta = R / m;
   group = zeros(m, 2);
   %указываем границы интервалов
   for j = 1:m
      group(j,1) = X(1)+delta*j;
   end
   %раскидываем элементы по интервалам
   j = 1;
   i = 1;
   border = X(1)+delta;
   while i < n
      if X(i) >= border && border < X(n) && j < 8
         border = border + delta;
         j = j + 1;
         continue; %чтобы корректно обрабатывать ситуациии, когда в
```

```
    %интервал не попадает ни один элемент выборки
    end
    group(j, 2) = group(j, 2) + 1;
    i = i + 1;
end
group(m, 2) = group(m, 2) + 1; %последний (n-й) элемент всегда будет
%принадлежать последнему (m-му) интервалу

for i = 1:m
    if i == 1
        fprintf('[%.6f', minX);
    else
        fprintf('[%.6f', group(i-1, 1));
    end
    fprintf(' - %.6f', group(i, 1));
    if i == m
        fprintf(']');
    else
        fprintf(')');
    end
    fprintf(': %d elements\n', group(i, 2));
end

% функция распределения (син)
% эмпирическая функция распределения
Xlen = zeros(1, n+2);
Xlen(1) = X(1) - 1;
for j = 1:n
    Xlen(j+1) = X(j);
end
Xlen(n+2) = X(n) + 1;

nn = length(Xlen);
Mmin = min(Xlen);
Mmax = max(Xlen);
step = (Mmax - Mmin) / nn;
xs = Mmin:step:Mmax;

%получаем функцию распределения для нормальной случайной величины
F = normcdf(xs, mu, sqrt(di));

%строим эмпирическую функцию распределения
E = zeros(nn, 1);
for i = 1:nn
    count = 0;
    for j = 1:n
        if X(j) <= Xlen(i)
            count = count+1;
        end
    end
    E(i) = count / n;
end
%строим графики
hold on;
plot(xs, F, "--"), grid;
stairs(Xlen, E), grid;
```

```
    figure();

    % гистограмма и функция плотности
    gist = zeros(m,2);
    gist(1,1) = (X(1) + group(1, 1))./ 2;
    for i = 2:m
        gist(i,1) = (group(i-1,1) + group(i, 1)) ./ 2;
    end
    %модифицируем Y для гистограммы - количество_попаданий / (n*delta)
    for i = 1:m
        y = group(i,2);
        y = y / (n*delta);
        gist(i,2) = y;
    end
    %вычисляем значения функции плотности распределения для всех X из выборки
    F = normpdf(X, mu, sqrt(di));
    %отображаем значения
    bar(gist(:,1), gist(:,2), 1);
    hold on;

    plot(X, F,'r'), grid;
end
```

5. Результаты расчетов для выборки

```
Min = 6.81
Max = 12.41
R = 5.60
mu = 9.49
sigma^2 = 1.21
m = 8

[6.810000 .. 7.510000): 7 elements
[7.510000 .. 8.210000): 4 elements
[8.210000 .. 8.910000): 21 elements
[8.910000 .. 9.610000): 36 elements
[9.610000 .. 10.310000): 27 elements
[10.310000 .. 11.010000): 14 elements
[11.010000 .. 11.710000): 7 elements
[11.710000 .. 12.410000]: 4 elements
```

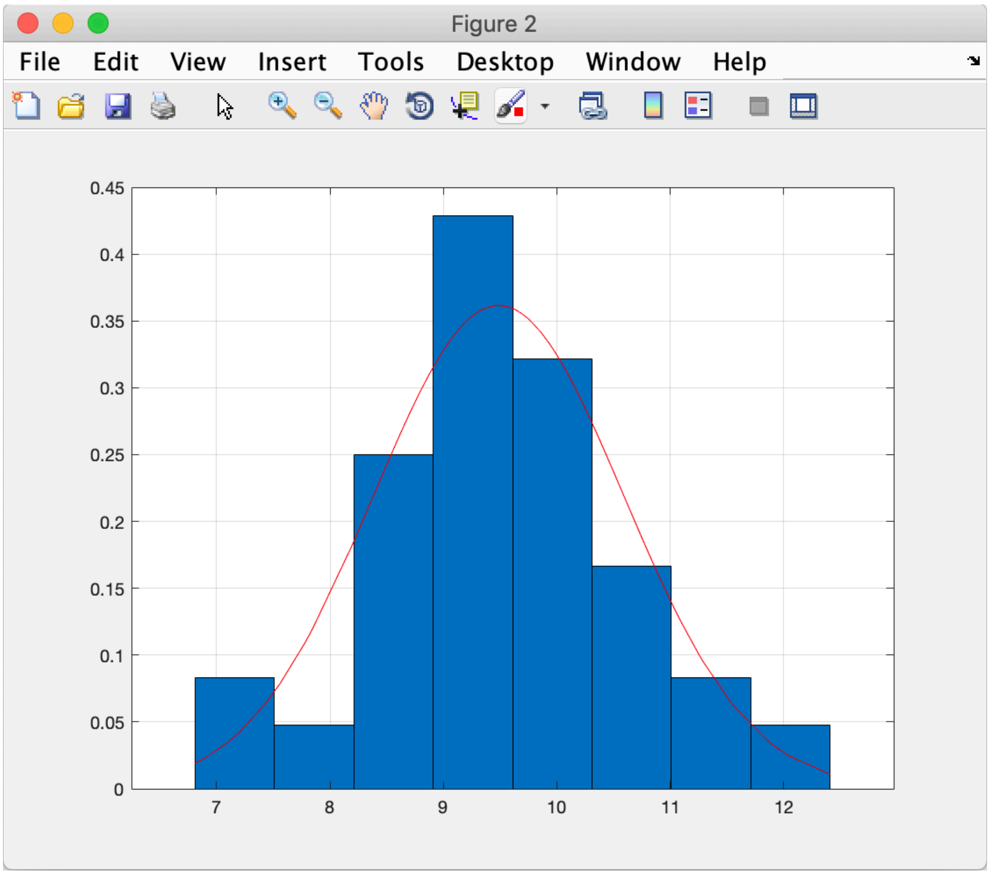

Рис.1 - Гистограмма и функция плотности

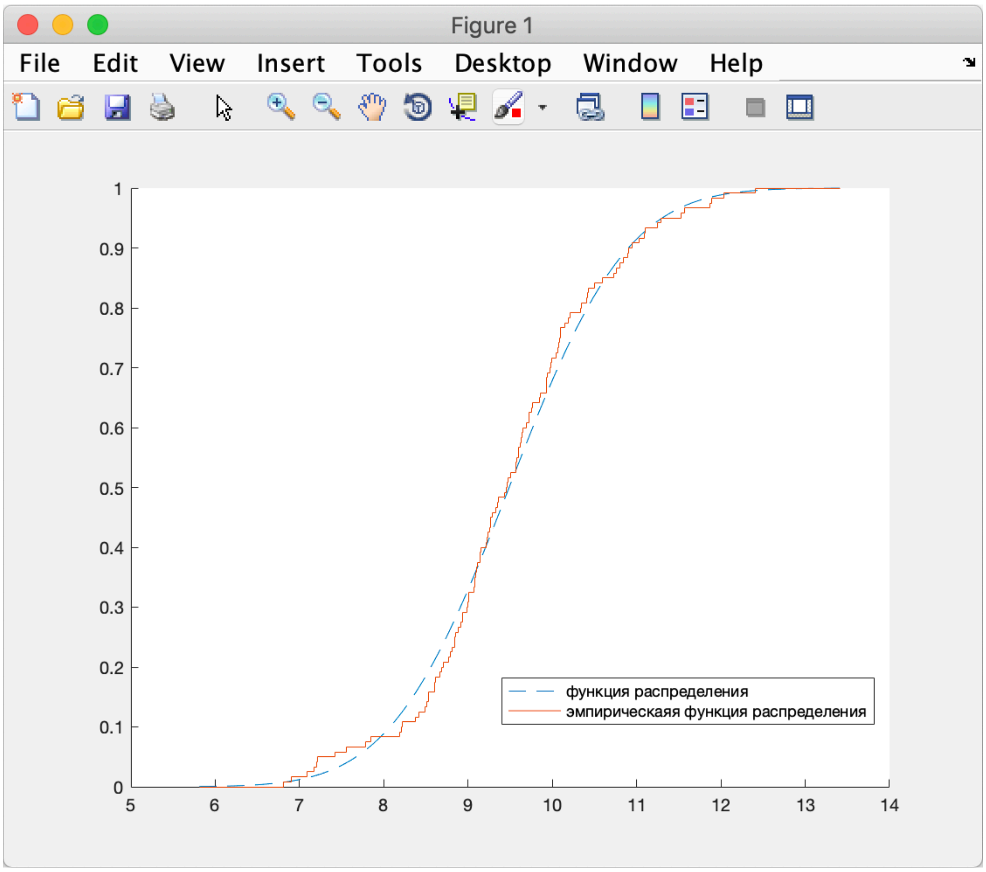

Рис.2 - Функция распределения,
эмпирическая функция распределения